Accepted Article

Genome-wide association study and meta-analysis on alcohol-related liver cirrhosis identifies novel genetic risk factors

Tae-Hwi Schwantes-An[1#], Rebecca Darlay[2#], Philippe Mathurin[3], Steven Masson[4], Suthat Liangpunsakul[5], Sebastian Mueller[6], Guruprasad P. Aithal[7], Florian Eyer[8], Dermot Gleeson[9], Andrew Thompson[10], Beat Muellhaupt[11], Felix Stickel[11], Michael Soyka[12], David Goldman[13], Tiebing Liang[5], Lawrence Lumeng[5ǂ], Munir Pirmohamed[10], Bertrand Nalpas[14,15], Jean-Marc Jacquet[14], Romain Moirand[16], Pierre Nahon[17-19], Sylvie Naveau[20], Pascal Perney[21], Greg Botwin[22,23], Paul S. Haber[24,25], Helmut K. Seitz[6], Christopher P. Day[4], Tatiana M. Foroud[1]*, Ann K. Daly[4]*, Heather J. Cordell[2]*, John B. Whitfield[26]*, Timothy R. Morgan[22]*[§], Devanshi Seth[24,25,27]*[§], for the GenomALC Consortium.

Keywords: *FAF2*, lipid droplet pathway, α1-antitrypsin, *PNPLA3*, *HSD17B13*

[1]Department of Medical and Molecular Genetics, Indiana University, USA. tlschwan@iu.edu, tforoud@iu.edu

[2]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, United Kingdom. rebecca.darlay@newcastle.ac.uk, heather.cordell@newcastle.ac.uk

[3]CHRU de Lille, Hôpital Claude Huriez, Rue M. Polonovski CS 70001, 59 037 Lille Cedex, France. Philippe.MATHURIN@CHRU-LILLE.FR

[4]Faculty of Medical Sciences, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom. Steven.Masson@nuth.nhs.uk, chris.day@newcastle.ac.uk, a.k.daly@newcastle.ac.uk

[5]Division of Gastroenterology and Hepatology, Department of Medicine, Indiana University, USA. sliangpu@iu.edu, tliang@iu.edu

[6]Department of Internal Medicine, Salem Medical Center and Center for Alcohol Research, University of Heidelberg, Zeppelinstraße 11e33, 69121 Heidelberg, Germany. sebastian.mueller@urz.uni-heidelberg.de, helmut_karl.seitz@urz.uni-heidelberg.de

[7]NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals and the University of Nottingham, Nottingham NG7 2UH, United Kingdom. Guru.Aithal@nottingham.ac.uk

[8]Division of Clinical Toxicology, Department of Internal Medicine 2, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany. florian.eyer@tum.de

Accepted Article

[9]The Clinical Research Facility, O Floor, The Royal Hallamshire Hospital, Glossop Road, S10 2JF, United Kingdom. Dermot.Gleeson@sth.nhs.uk

[10]MRC Centre for Drug Safety Science, Liverpool Centre for Alcohol Research, University of Liverpool, The Royal Liverpool and Broadgreen University Hospitals NHS Trust, and Liverpool Health Partners, Liverpool, L69 3GL, UK. Andrew.Thompson@liverpool.ac.uk, munirp@liverpool.ac.uk

[11]Department of Gastroenterology and Hepatology, University Hospital Zurich, Rämistrasse 100, CH-8901 Zurich, Switzerland. beat.muellhaupt@usz.ch, Felix.Stickel@hirslanden.ch

[12]Psychiatric Hospital University of Munich, Nussbaumsstr.7, 80336 Munich, Germany and Privatklinik Meiringen, Willigen, CH 3860 Meiringen, Switzerland. michael.soyka@privatklinik-meiringen.ch

[13]Laboratory of Neurogenetics, NIAAA, Rockville, MD 20852, USA. davidgoldman@mail.nih.gov

[14]Service Addictologie, CHRU Caremeau, 30029 Nîmes, France. bertrand.nalpas@inserm.fr, jeanmarc.jacquet@chu-nimes.fr

[15]DISC, Inserm, 75013 Paris, France.

[16]Univ Rennes, INRAE, INSERM, CHU Rennes, Institut NUMECAN (Nutrition Metabolisms and Cancer), F-35000 Rennes, France. romain.moirand@univ-rennes1.fr

[17]APHP, Liver Unit, Hospital Jean Verdier, Bondy, France. pierre.nahon@aphp.fr

[18]University Paris 13, Bobigny, France.

[19]Inserm U1162 “Functional Genomics of Solid Tumors”, Paris, France.

Accepted Article

[20]Hôpital Antoine-Béclère, 157 Rue de la Porte de Trivaux, 92140 Clamart, France. sylvienaveau@orange.fr

[21]Hôpital Universitaire Car_emeau, Place du Pr. Robert Debr_e, 30029 Nîmes, France. Pascal.PERNEY@chu-nimes.fr

[22]Medical and Research Services, VA Long Beach Healthcare System, 5901 East Seventh Street, Long Beach, CA 90822, USA. timothy.morgan@va.gov

[23]Translational Genomics Group, Inflammatory Bowel & Immunobiology Research Institute, 8700 Beverly Blvd., Thalians Bldg. #E244: Los Angeles CA 90048. gregory.botwin@cshs.org

[24]Drug Health Services, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia. paul.haber@sydney.edu.au

[25]Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2006, Australia.

[26]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Queensland 4029, Australia. John.Whitfield@qimrberghofer.edu.au

[27]Centenary Institute of Cancer Medicine and Cell Biology, The University of Sydney, Sydney, NSW 2006, Australia. d.seth@sydney.edu.au

ǂ Dr Lumeng passed away on 21st June 2017

Equal first authors

* Equal senior authors

§ Co-corresponding authors

Accepted Article

[§]Devanshi Seth, PhD. Centenary Institute, 93 Missenden Road, Camperdown, NSW 2050, Australia. Phone: +61-2-9565-6268 Fax: +61-2 9565-6101 Email: d.seth@sydney.edu.au

[§]Timothy R. Morgan, MD. Medical Service -11, VA Long Beach Healthcare System, 5901 East Seventh Street, Long Beach, CA 90822, USA. Phone: +1 562 826 5756
Fax: +1-562 826 5436 Email: timothy.morgan@va.gov

Abbreviations: AAT, α-1-antitypsin; ALT, Alanine transaminase; ALC, Alcohol-related liver cirrhosis; AST, Aspartate transaminase; ATGL, Adipose triglyceride lipase; BMI, Body Mass Index; CI, 95% Confidence Interval; FAF2, Fas Associated Factor Family Member 2; GWAS, Genome-wide Association Study; HCC, hepatocellular carcinoma; HSD17B13, 17-β-hydroxysteroid dehydrogenase type 13; HWE, Hardy-Weinberg Equilibrium; ICD, International Statistical Classification of Diseases and Related Health Problems; LD, Linkage Disequilibrium; LNG, Laboratory of Neurogenetics; MAF, Minor Allele Frequency; NASH, Non-alcoholic Steatohepatitis; NAFLD, Non-alcoholic fatty liver disease; PC, Principal Components; OR, Odds Ratio; PNPLA3, Patatin Like Phospholipase Domain Containing 2; SNP, single nucleotide polymorphism; T2DM, type 2 diabetes mellitus

FINANCIAL SUPPORT STATEMENT

This investigation was funded by the National Institutes of Health and National Institute on Alcohol Abuse and Alcoholism U01- AA018389. The funds were used for recruitment of participants, including researcher/nurse time, data and sample collection, and genotyping. Further support came from grants from the Swiss National Funds (SNF no. 310030_169196) and the Swiss Foundation for Alcohol Research (SSA) for FS, and from National Health & Medical

Accepted Article

Research Council (NHMRC)/Medical Research Future Funds (MRFF) Practitioner Fellowship (APP1155320) for PH. We confirm the independence of researchers from funders and that all authors, external and internal, had full access to all the data (including statistical reports and tables) in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Accepted Article

Accepted Article

Abstract

Only a minority of heavy drinkers progress to alcohol-related cirrhosis (ALC). The aim of this study was to identify common genetic variants that underlie risk for ALC. We analyzed data from 1,128 subjects of European ancestry with ALC and 614 heavy drinking subjects without known liver disease from Australia, the United States, the United Kingdom and three countries in Europe. Genome-wide association study (GWAS) was performed, adjusting for principal components and clinical covariates (alcohol use, age, sex, BMI, diabetes). We validated our GWAS findings using UK Biobank. We then performed a meta-analysis combining data from our study, the UK Biobank, and a previously published GWAS. Our GWAS found genome-wide significant risk association of rs738409 in *PNPLA3* (Odds Ratio (OR)=2.19 (G allele), p-value=4.93×10^{-17}) and rs4607179 near *HSD17B13* (OR=0.57 (C allele), p-value=1.09×10^{-10}) with ALC. Conditional analysis accounting for the *PNPLA3* and *HSD17B13* loci identified a new protective association at rs374702773 in *Fas Associated Factor family member 2* (*FAF2*) (OR=0.61 (del(T) allele), p-value=2.56×10^{-8}) for ALC. This association was replicated in the UK Biobank using conditional analysis (OR=0.79, p-value=0.001). Meta-analysis (without conditioning) confirmed genome-wide significance for the newly identified *FAF2* locus as well as *PNPLA3* and *HSD17B13*. Two other previously known loci (*SERPINA1, SUGP1/TM6SF2*) were also genome-wide significant in the meta-analysis. GeneOntology pathway analysis identified lipid droplets as the target for several identified genes. In conclusion, our GWAS identified a new locus at *FAF2* associated with reduced risk of ALC among heavy drinkers. Like the *PNPLA3* and *HSD17B13* gene products, the *FAF2* product has been localized to fat droplets in hepatocytes. Our genetic findings implicate lipid droplets in the biological pathway(s) underlying ALC.

Accepted Article

Chronic alcohol use is a leading cause of cirrhosis in the Western world and is on the rise in many other countries. Although chronic and heavy alcohol use is a requirement for development of alcohol-related liver cirrhosis (ALC), only a minority of heavy drinkers progress to cirrhosis. Reliable risk estimates for cirrhosis among heavy drinkers are difficult to obtain, but it is often estimated at 10-15% after decades of heavy alcohol use (1, 2). Generally accepted risk factors for the development of ALC include duration and amount of alcohol consumed, female gender, and obesity.

Many authors have suggested genetic variation in vulnerability to alcohol-related, or other types of liver disease. The most reported genetic variant in liver diseases is rs738409 in *PNPLA3*. The variant's G allele is associated with increased risk for non-alcoholic fatty liver disease/steatohepatitis (NAFLD/NASH) (3) and with increased risk and severity for alcohol-related liver diseases (4-6). A genome-wide association study (GWAS) of ALC, from two cohorts comprising approximately 700 alcohol-related cirrhotic cases and 1400 drinking controls without known liver disease, reported a genome-wide-significant association between rs738409 and ALC (7). These investigators also reported association of rs58542926 (*TM6SF2*), and rs641738 (*MBOAT7*) with ALC (7). Additionally, Abul-Husn et al. reported an association between rs72613567 (*HSD17B13*) and alcohol-related and non-alcoholic fatty liver-related cirrhosis (8), although not at a genome-wide significance level. Most recently, Emdin et al. reported an association between rs2642438 (*MARC1*) and all-cause cirrhosis, including in a subset of patients with alcohol-related liver diseases (9). Thus, several analyses have identified Single Nucleotide Polymorphisms (SNPs) associated with both alcohol-related liver disease and

NASH, supporting a genetic predisposition for ALC, as well as shared genetic susceptibility with NASH (10).

We undertook a multinational GWAS of European ancestry participants with carefully characterized ALC status to identify predisposing genetic factors. We enrolled subjects at high and low ends of the spectrum of alcohol-related liver disease, specifically excluding subjects with intermediate stages of disease to maximize statistical power. Cases had a history of high-risk alcohol intake with clinically or histologically defined ALC. Controls were heavy drinkers without clinical evidence of liver disease. We chose drinkers without known liver disease rather than non-drinkers as controls to avoid the potential of finding genetic risk loci related to alcohol use. Our stringent eligibility criteria for this case-control design allowed for standardization of both the ALC subjects and of the non-liver disease subjects across enrollment sites.

Material and Methods

SUBJECTS

GenomALC cohort

The GenomALC study participants were recruited at clinical sites in Australia, France, Germany, Switzerland, the United Kingdom, and the United States using a standardized pre-defined protocol (11). Enrollment occurred between 2012 and 2017 with approvals from site-specific governing ethics committees and written informed consent from all recruited participants. All participants were required to have alcohol consumption of ≥80 grams per day (males) and ≥50 grams per day (females) for at least 10 years. Controls were defined as having normal bilirubin, AST, and ALT levels at the time of heavy alcohol use and no prior evidence of liver injury.

Accepted Article

Accepted Article

Those with mildly elevated liver tests were included as controls if their transient elastography was less than 6 kPa. Cases were defined by 1) clinically evident portal hypertension or decompensated cirrhosis (e.g., ascites, esophageal varices), 2) Fibroscan stiffness >22kPa if AST <100 IU/L/>32kPa if AST 100-200 IU/L/exclude if AST >200 IU/L, or 3) liver histology (Metavir score of F4) on a previously-performed liver biopsy. Those with other causes of liver disease (e.g. viral hepatitis, hemochromatosis) or with HIV were excluded as described in detail previously (11). DNA was obtained from blood samples and genotyped on Global Screening Array v1.0 (Illumina., 5200 Illumina Way, San Diego, CA 92122). Genotype data were reviewed using the steps detailed in the Supplemental texts. In brief, using a GWAS data cleaning pipeline, genotyped SNPs were filtered for call rate, violation of the Hardy-Weinberg equilibrium (HWE) and minor allele frequency (MAF). Samples were checked for genotyping rate, sex, relatedness, and European genetic ancestry. PLINK was used to check for discordance between reported and genetic sex, sample relationship was calculated using --genome command and using Pi-hat value, we removed one sample from each pair of related individuals down to second-degree relatives. Genetic ancestry was determined using SNPRelate package using 1000 Genomes Project as reference. Additional genotypes were imputed using the Michigan Imputation Server (12). Additional details on GWAS data cleaning is available in the Supplement Texts under 'Genetic Data Cleaning' section.

Laboratory of Neurogenetics (LNG) cohort

Genotype and phenotype data from 860 heavy-drinking subjects who had given consent as part of two NIH screening/natural history protocols (98-AA-009; 05-AA-0121) were provided by the Laboratory for Neurogenetics (LNG) at the National Institutes on Alcohol Abuse and Alcoholism (NIAAA, Rockville, MD, USA) as additional heavy drinking controls. Participants

were required to have alcohol consumption of ≥80 grams per day (males) and ≥50 grams per day (females) and excluded if they had a serious medical condition requiring ongoing treatment, including liver disease. Genotype data (Illumina OmniExpress Bead Chip) were cleaned and imputed using the same pipeline used for GenomALC. After filtering, 235 participants meeting the inclusion criteria were included as controls.

UK Biobank cohort

Data from the UK Biobank (https://www.ukbiobank.ac.uk/) were accessed under approval number 18870. Cirrhosis cases (n=530) were defined as having ICD10 code K70.3; 'Alcoholic Cirrhosis of Liver', or ICD-9 code 571.2, 'Cirrhosis, liver, alcoholic', and ICD10 code K70.1; 'Alcoholic hepatitis without ascites' or ICD-9 code 571.1, 'Acute alcoholic hepatitis'. Controls (n=10,222) were defined as having 1) reported alcohol intake of ≥80g/d (males) and ≥50g/d (females) and/or 2) an ICD10 diagnosis of F10 (Mental and behavioural disorders due to alcohol), but with no recorded diagnosis of any liver disease. Genotype data were cleaned and imputed by the UK Biobank investigators; we used the 'White British' ancestry subset of samples provided. After additional data cleaning (see Supplemental Text), 8,185,141 SNPs in 439 alcohol-related cirrhosis cases and 8,364 controls were available for analysis.

Additional sample description and genotype data cleaning information for the three cohorts is available in the Supplemental Texts.

Accepted Article

Accepted Article

STATISTICAL METHODS

Genome Wide Association Tests in GenomALC

A total of 1,128 ALC cases and 614 heavy drinking controls were available for analysis. To test association between each SNP and ALC (cases vs. controls), logistic regression between case-control status and dosage value of each SNP was tested using PLINK (13). The following covariates were included in each regression model: 1) first ten principal components (PCs), 2) age, 3) sex, 4) years of excessive drinking, 5) alcohol consumed measured in grams per day, 6) total lifetime alcohol consumed measured in kilograms, 7) diabetes status (present/absent), and 8) body mass index (BMI). To identify additional genetic associations beyond the two genome-wide significant loci (rs4607179 for *HSD17B13* and rs738409 for *PNPLA3*), we ran a conditional GWAS that included minor allele counts (dosage) for variants at the two loci as covariates. Conditional GWAS also included all covariates used in the primary GWAS along with the two dosage values. The p-value threshold for genome-wide statistical significance was set at 5×10^{-8}.

GWAS in UK Biobank data

The data were analyzed with FaST-LMM (14), a mixed model method that computes a kinship matrix to adjust for relatedness/population stratification and infers any missing genotypes at a SNP based on the genotypes of other samples at that SNP. Cases (231) and drinking controls (8,364) that had all the clinical covariates (age, sex, type 2 diabetes mellitus (T2DM) status, BMI, and current daily alcohol intake (in grams)) were included in a GWAS that mirrored the primary analysis performed in GenomALC cohort. To mirror the GenomALC conditional analysis in the UK Biobank data, cases and drinking controls used above were re-analyzed with

the minor allele counts at rs738409 and rs4607179 as covariates along with the clinical variables. ORs were obtained from logistic regression in PLINK including all relevant covariates, and confidence intervals were calculated from back-transformation of FaST-LMM p-values and PLINK ORs.

As a sensitivity analysis, we also conducted a GWAS of cases and controls that excluded those that did not meet the criteria of current alcohol intake per day ≥ 80g/day (males) and ≥ 50g/day (females), resembling the GenomALC criteria. This GWAS included a total of 231 cases and 5,361 controls and adjusted for the same set of covariates used in the primary GWAS.

Meta-analysis

Meta-analysis of our primary GWAS results from GenomALC, the UK Biobank, and the GWAS summary statistics from a published GWAS of ALC (7) was conducted using METAL (15) with genomic control option. Summary statistics for the published GWAS (7) were obtained from http://gengastro.med.tu-dresden.de/suppl/alc_cirrhosis/.

Gene-set enrichment analysis

We used enrichment analysis (16) from Gene Ontology (17, 18) to identify potential pathways shared by genes identified in our GenomALC GWAS analysis. Using the genes that includes or are nearby the SNPs with p-values less than $5x10^{-6}$ from our primary GWAS analysis, we identified 15 genes for enrichment analysis. We performed enrichment analysis for molecular function, cellular component, and biological processes from Gene Ontology (17, 18).

Accepted Article

Accepted Article

Results

PRIMARY GWAS ANALYSIS

The primary GenomALC GWAS analysis identified two genome-wide significant associations (Figure 1, Supplementary Figure S1). rs738409 on chromosome 22 (chr22), located in *PNPLA3,* was significantly associated with increased risk of ALC, with a p-value of 4.93×10^{-17} and an OR of 2.19 (95% confidence interval (CI), 1.81 - 2.54) for each copy of the G allele. rs4607179 on chr4, located near *HSD17B13,* was associated with lower risk of ALC with a p-value of 1.09×10^{-10} and OR of 0.57 for each copy of the C allele (95% CI, 0.48 - 0.62) (Table 1). A list of all SNPs with p-values less than 5×10^{-6} and their association test results are shown in Supplementary Table S1. Age, sex, and 10 PCs adjusted GWAS analysis result of GenomALC and additional controls from LNG is shown in Supplementary Figure S2 and Table S2. Descriptive statistics of the GenomALC cohort are shown in Table 2 and UK Biobank and LNG are shown in Table S3. Primary GWAS analysis results from UK Biobank and sensitivity analysis, performed on controls reporting current alcohol use >80/50 grams/day are shown in Supplementary Tables S4 and S5, respectively.

CONDITIONAL GWAS ANALYSIS

To identify additional genetic associations, we performed a conditional GWAS that included minor allele counts (dosage) for variants in the two genome-wide significant loci (rs4607179 for *HSD17B13* and rs738409 for *PNPLA3*) from the primary GWAS analysis as covariates along with the previously included covariates. The conditional GWAS analysis identified genome-wide significant association for rs374702773 (*FAF2*) on chr5 with a p-value of 2.56×10^{-8} and an OR of 0.61 (95% CI, 0.51 - 0.73) for each copy of the 7bp deletion allele (Figure 2, Supplementary

Figure S3, Table 1). Association test results for both primary and conditional GWAS analyses for the *FAF2* locus from GenomALC are shown in Supplementary Table S6. Figure 3 shows a LocusZoom (19) plot of rs374702773 and nearby variants. rs374702773 was not available in the UK Biobank data, instead we tested four variants in strong LD (D'≥0.9 & r^2≥0.6) with it. Linkage disequilibrium (LD) values among variants are shown in Supplementary Table S7. In the conditional analysis of the UK Biobank, rs11134977 showed the strongest association with ALC (OR=0.79, p-value of 0.001), other SNPs (rs11027, rs12514451, rs34152523) in LD with rs374702773 showed similar OR and p-value. (Supplementary Table S6). We tested for interaction between *FAF2* locus (rs374702773 and 4 SNPs in LD) and *PNPLA3* (rs738409) and *FAF2* locus and *HSD17B13* (rs4607179) in GenomALC. None of the interaction terms were statistically significant (p-values >0.44) (Table S8).

META-GWAS OF THREE STUDIES

Meta-analysis of results from GenomALC, UK Biobank, and a previously published GWAS of alcohol-related liver cirrhosis (7) identified five independent genome-wide significant associations with ALC (Figure 4, Supplementary Figure S4). Summaries of top associated SNPs from each locus are shown in Table 1 and all SNPs with p-values less than $5x10^{-6}$ are shown in Supplementary Table S9. In the meta-analysis, the most significant association at *FAF2* was for rs11134977 (p-value = 1.56×10^{-8}), which is in LD with rs374702773 in *FAF2*. In addition to associations at the *HSD17B13*, *PNPLA3*, and *FAF2* loci, we identified two additional genome-wide significant associations, rs28929474 (chr14:94,844,947, *SERPINA1*, OR 1.90, 95% C.I. 1.52-2.38, p-value=$1.99x10^{-8}$) and rs10401969 (chr19:19,407,718, *SUGP1*, OR 1.49 95% C.I. 1.31-1.70, p-value $2.40x10^{-9}$) (Table 1). LocusZoom plots of these five loci are available in

Accepted Article

Supplementary Figure S5. Meta-analysis of Age, sex, and 10 PCs adjusted GWAS in GenomALC+LNG and UK Biobank and published GWAS from Buch et al. cohorts is available in Supplementary Figure S6 and Table S10.

GENE ENRICHMENT ANALYSIS

Gene enrichment analysis for biological process and cellular component showed a significant enrichment for lipid droplet and its organization among the 15 genes (Supplementary Table S11). Enrichment for biological process was observed for lipid droplet organization (>100-fold, FDR=0.003) and cellular component (>70-fold, FDR=0.020) (Table 3).

Discussion

This is the largest-to-date GWAS for alcohol-related cirrhosis, and the only multi-center GWAS of prospectively enrolled subjects meeting strict, protocol-defined criteria for heavy drinking, and for alcohol-related cirrhosis or no known liver disease. We also conducted the first GWAS meta-analysis in ALC contrasting clinically defined groups of cases and heavy drinking controls from GenomALC, the UK Biobank, and a previously published GWAS on ALC (7). We report a new genome-wide association with ALC; rs374702773 located in *Fas Associated Factor family member 2 (FAF2),* with the C (minor) allele showing a protective effect. Additionally, we replicated the association between rs738409 (*PNPLA3*) and ALC (7) strengthening previous reports (8, 20). We also replicated an association between *HSD17B13* locus and ALC by identifying a genome-wide significant association between ALC and rs4607179 (p-value=1.09×10^{-10}). Meta-analysis also identified two additional genome-wide significant associations; rs28929474, a missense variant in *SERPINA1* and rs10401969 located in a linkage

Accepted Article

disequilibrium block that spans multiple genes including *SUGP1* and *TM6SF2*. However, we did not find genome-wide significance for rs641738 or any other SNPs in *MBOAT7* (primary GWAS p-value=$5.31x10^{-03}$, meta-analysis p-value=$4.03x10^{-05}$) which was reported by Buch et al (7). A recently reported association of rs2642438 in *MARC1* gene and cirrhosis (9) was also not genome-wide significant in our meta-analysis (OR=0.81, p-value=$7.54x10^{-06}$) nor in GenomALC GWAS (OR=0.79, p-value=0.006), but the effect was in the same direction.

FAF2 represents a novel locus affecting genetic risk for ALC. This association became genome-wide significant for rs374702773 in our GenomALC cohort after accounting for the effects of rs738409 (*PNPLA3*) and rs4607179 (near *HSD17B13*) (Supplementary Table S6). We also replicated this association using UK Biobank cohort (OR=0.79, p-value=0.001 for rs11134977). Using the larger sample size available in the meta-analysis of GenomALC, UK Biobank, and Buch et al (7), this association was found without conditional analysis (OR=0.79, p-value=$1.56x10^{-08}$ for rs11134977). We did not observe interaction between *FAF2* locus and *PNPLA3* or *HSD17B13* loci (Supplementary Table S8). Although this may indicate that these loci act independently, our study was not powered enough to detect interactions. These variants in *FAF2* are in strong LD and represent the same association. Also known as UBDX8, FAF2 is a sensor of intracellular levels of long-chain unsaturated fatty acids (21). In the presence of unsaturated fatty acid, UBXD8 binds to adipose triglyceride lipase (ATGL), the major triglyceride hydrolyzing enzyme, and dissociates it from its activator α/β hydrolase domain 5 (ABHD5, also known as comparative gene identification-58 [CGI-58]). This dissociation from ABHD5/CGI-58 inhibits triacylglycerol hydrolysis in lipid droplets (21, 22). This biochemical pathway, that is, binding to ABDH5/CGI-58 and inhibiting triglyceride hydrolysis in lipid droplets, is believed to be one mechanism by which *PNPLA3* (I148M) contributes to hepatic

Accepted Article

steatosis (23, 24). Thus, *FAF2* fits into the current proposed genetic pathways that affect hepatocyte lipid metabolism. However, *FAF2* also regulates SREBP-1 activation and triacylglycerol synthesis from diacylglycerol (DAG) by DAG acytltransferase (21). The role and molecular function of the *FAF2* associated variant and those SNPs in LD with it, are yet to be explored. Thus, the actual role of *FAF2* in ALC risk will require a better understanding of the pathophysiology of liver injury by *FAF2* and other implicated genes.

Meta-analysis found genome-wide significance for association between rs28929474, a missense variant of *SERPINA1*, and ALC. *SERPINA1* codes for the α-1-antitypsin (AAT) protein, with rs28929474 (Glu366His) coding for the Z variant of α-1-antitypsin. The Z variant causes conformational rearrangement of α1-antitrypsin, affecting its degradation by proteases (25, 26). Homozygosity for the Z variant of *SERPINA1* leads to AAT deficiency (27, 28) and increased risk for cirrhosis. Although this is a variant with low MAF, there is independent evidence that it contributes to the risk of alcohol-related liver disease and cirrhosis. Heterozygosity of this variant has also been associated with portal hypertension (29) and cystic fibrosis-related liver disease (30). Abul-Husn et al. (8) and Chen et al. (31) reported rs28929474 association with alcohol-related liver disease/cirrhosis and non-alcoholic liver diseases using candidate variant analysis and a GWAS level significance with liver transaminases. Strnad and colleagues reported an association between the Z variant heterozygous carriers and cirrhosis in NAFLD (OR=7.3, $p<0.0001$) and in alcohol misusers (OR=5.8, $p<0.0001$) (32). Although none of these associations approached GWAS-level significance, their findings are in the same direction as our report. Taken together, this is a new disease risk associated with the AAT Z-variant gene product as up to now it was thought that AAT-related liver disease only occurs in ZZ homozygotes.

Accepted Article

We extended an association between *HSD17B13* (17-β-hydroxysteroid dehydrogenase type 13) locus and liver disease reported previously. Although not at genome-wide significance level, these authors reported that the A allele (minor allele) of rs72613567 showed protective effect against NAFLD and/or alcohol-related liver diseases including cirrhosis and hepatocellular carcinoma (8). However, most recently, another group reported GWAS significance for *HSD17B13* associated with protection against NAFLD (33). In our discovery GWAS of GenomALC we found genome-wide significance for rs4607179 near *HSD17B13*. Our meta-analysis confirmed significance for SNPs in *HSD17B13* and risk for ALC; we observe protective effect of the C allele of rs4607179 on ALC at genome-wide significance level (OR=0.80, $p=1.39x10^{-08}$). Although the detailed biological function of *HSD17B13* is not fully understood, it too is known to be associated with hepatic lipid droplets (34) and circulating triglyceride and high density lipoprotein levels (35, 36), suggesting overlapping mechanisms in ALC and other complex liver diseases.

We replicated the association between ALC and rs738409, which codes for a missense mutation in *PNPLA3*. rs738409 had the strongest statistical association and the highest OR among the genome-wide significant SNPs (OR=2.19, p-value=$4.93x10^{-17}$). This locus has been associated with risk for NAFLD, alcohol-related liver disease, and HCC in several meta-analyses (37-40). PNPLA3 is a lipase found on the surface of triglyceride droplets in hepatocytes. The variant PNPLA3 protein may increase hepatic steatosis by increasing dissociation of ABDH5/CGI-58 from adipose triglyceride lipase, thereby inhibiting the triglyceride catabolism by ATGL (23). In addition, the half-life of the rs738409 variant PNPLA3 protein on the lipid droplet appears prolonged, extending its ability to inhibit triglyceride hydrolysis (23, 24).

Accepted Article

Our meta-analysis found genome-wide significance for rs10401969 near *SUGP1* on chr19 (OR=1.49 for C allele, p-value=2.40×10^{-09}). rs1041969 has been proposed to modulate activity of 3-hydroxy-3-methylglutaryl-CoA reductase (also known as HMG-CoA reductase), the rate limiting enzyme for cholesterol synthesis. rs10401969 has been reported to show associations with plasma LDL cholesterol level, coronary artery disease, hepatic fat in obese patients (41), and elevated ALT in heavy alcohol drinkers (42), suggesting that *SUGP1* may be directly related to genetic risk for ALC. However, *SUGP1* is part of a locus with multiple genes and this intron variant is in strong LD (r^2 ~0.95 in Europeans) with variants located in nearby *TM6SF2* (Supplementary Figure S5), a locus which has been associated with ALC (7), NAFLD (43-45), cirrhosis/steatosis/fibrosis among chronic hepatitis C patients (46), and impaired lipid synthesis (47). Thus, it is not clear whether this variant implicates *SUGP1* as a potential underlying gene for risk of cirrhosis or is a marker for *TM6SF2*.

One common thread among our new and replication findings is the involvement of genetic variants in lipid biology, particularly triglyceride metabolism. Enrichment analysis using GeneOntology (17) identified lipid droplet cellular components and lipid droplet organization biological processes among the genes identified by GWAS. The genes involved in the process from our list of candidate genes were *HSD17B13*, *PNPLA3*, and *FAF2*. However, these genes have pleotropic effects: FAF2 also regulates SREBP-1 activation and triacylglycerol synthesis (21) while HSD17B13 has retinol dehydrogenase activity (25). Thus, defining the actual role of these genes in ALC will depend on a better understanding of lipid droplet biology, and other biochemical pathways, in the pathophysiology of liver injury especially for non-alcohol related disease, which shares several underlying genetic factors and pathophysiologic mechanisms (48, 49).

Accepted Article

This GWAS has potential limitations. Despite being the largest available cohort for ALC, the size of our sample for a GWAS of common variants imposes a limit on the power to detect associations with weaker effect sizes. Second, although the trend for the odds ratios in the UK Biobank cohort is in the same direction as for GenomALC, the strength of the association was weaker. We suspect the weaker association may be due to the smaller number of cases in UK Biobank and expanded definition of heavy drinking controls in the UK Biobank as compared with GenomALC. Using stricter drinking criteria for controls, in our secondary and sensitivity analysis in the smaller UK Biobank cohort, we observed similar association results (Figure S8).

Overall, we report a new genetic association between *FAF2* and ALC. This locus is functionally linked to two previously reported loci (*PNPNA3* and *HSD17B13*) by a biological pathway that governs lipid droplet organization. We also report increased risk for ALC in heterozygous carriers of the *SERPINA1* Z variant. Taken together, our findings suggest that risk of developing ALC is, in part, related to genetic factors, especially genetic factors regulating lipid homeostasis, and heavy alcohol exposure is a necessary but not sufficient pre-condition for ALC.

Accepted Article

Accepted Article

REFERENCES

1. Mann RE, Smart RG, Govoni R. The epidemiology of alcoholic liver disease. Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism 2003;27:209-219.

2. Askgaard G, Kjær MS, Tolstrup JS. Opportunities to Prevent Alcoholic Liver Cirrhosis in High-Risk Populations: A Systematic Review With Meta-Analysis. American Journal of Gastroenterology 2019;114:221-232.

3. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, Boerwinkle E, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nature Genetics 2008;40:1461-1465.

4. Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. Nat Genet 2010;42:21-23.

5. Seth D, Daly AK, Haber PS, Day CP. Patatin-like phospholipase domain containing 3: a case in point linking genetic susceptibility for alcoholic and nonalcoholic liver disease. Hepatology 2010;51:1463-1465.

6. Salameh H, Raff E, Erwin A, Seth D, Nischalke HD, Falleti E, Burza MA, et al. PNPLA3 Gene Polymorphism Is Associated With Predisposition to and Severity of Alcoholic Liver Disease. Am J Gastroenterol 2015;110:846-856.

7. Buch S, Stickel F, Trépo E, Way M, Herrmann A, Nischalke HD, Brosch M, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. Nature Genetics 2015;47:1443.

Accepted Article

8. Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, Liu Y, et al. A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. The New England Journal of Medicine 2018;378:1096-1106.

9. Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, Hindy G, et al. A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. PLOS Genetics 2020;16:e1008629.

10. Scott E, Anstee QM. Genetics of alcoholic liver disease and non-alcoholic steatohepatitis. Clinical medicine (London, England) 2018;18:s54-s59.

11. Whitfield JB, Rahman K, Haber PS, Day CP, Masson S, Daly AK, Cordell HJ, et al. Brief Report: Genetics of Alcoholic Cirrhosis—GenomALC Multinational Study. Alcohol Clin Exp Res 2015;39:836-842.

12. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, et al. Next-generation genotype imputation service and methods. Nature Genetics 2016;48:1284.

13. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4.

14. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. Nature Methods 2011;8:833-835.

15. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191.

16. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 2017;45:D183-d189.

Accepted Article

17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 2000;25:25-29.

18. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research 2018;47:D330-D338.

19. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England) 2010;26:2336-2337.

20. Stickel F, Lutz P, Buch S, Nischalke HD, Silva I, Rausch V, Fischer J, et al. Genetic variation in HSD17B13 reduces the risk of developing cirrhosis and hepatocellular carcinoma in alcohol misusers. Hepatology 2019;0.

21. Kim H, Ye J. Cellular responses to excess fatty acids: focus on ubiquitin regulatory X domain-containing protein 8. Curr Opin Lipidol 2014;25:118-124.

22. Olzmann JA, Richter CM, Kopito RR. Spatial regulation of UBXD8 and p97/VCP controls ATGL-mediated lipid droplet turnover. Proc Natl Acad Sci USA 2013;110:1345-1350.

23. Wang Y, Kory N, BasuRay S, Cohen JC, Hobbs HH. PNPLA3, CGI-58, and Inhibition of Hepatic Triglyceride Hydrolysis in Mice. Hepatology 2019;69:2427-2441.

24. **Yang A**, **Mottillo EP**, Mladenovic-Lucas L, Zhou L, Granneman JG. Dynamic interactions of ABHD5 with PNPLA3 regulate triacylglycerol metabolism in brown adipocytes. Nat Metab 2019;1:560-569.

25. Lomas DA, Li-Evans D, Finch JT, Carrell RW. The mechanism of Z α1-antitrypsin accumulation in the liver. Nature 1992;357:605-607.

26. Kopito RR, Ron D. Conformational disease. Nat Cell Biol 2000;2:E207-209.

Accepted Article

27. de Serres FJ, Blanco I, Fernandez-Bustillo E. PI S and PI Z alpha-1 antitrypsin deficiency worldwide. A review of existing genetic epidemiological data. Monaldi Arch Chest Dis 2007;67:184-208.

28. Zorzetto M, Russi E, Senn O, Imboden M, Ferrarotti I, Tinelli C, Campo I, et al. SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. Clin Chem 2008;54:1331-1338.

29. Mandorfer M, Bucsics T, Hutya V, Schmid-Scherzer K, Schaefer B, Zoller H, Ferlitsch A, et al. Liver disease in adults with alpha1-antitrypsin deficiency. United European Gastroenterol J 2018;6:710-718.

30. Boëlle P-Y, Debray D, Guillot L, Corvol H, on behalf of the French CFMGSI. SERPINA1 Z allele is associated with cystic fibrosis liver disease. Genetics in Medicine 2019;21:2151-2155.

31. Chen VL, Chen Y, Du X, Handelman SK, Speliotes EK. Genetic variants that associate with cirrhosis have pleiotropic effects on human traits. Liver International 2020;40:405-415.

32. **Strnad P**, **Buch S**, Hamesch K, Fischer J, Rosendahl J, Schmelz R, Brueckner S, et al. Heterozygous carriage of the alpha1-antitrypsin Pi*Z variant increases the risk to develop liver cirrhosis. Gut 2019;68:1099-1107.

33. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, Burt AD, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically-characterised cohort. J Hepatol 2020.

34. Su W, Mao Z, Liu Y, Zhang X, Zhang W, Gustafsson J-A, Guan Y. Role of HSD17B13 in the liver physiology and pathophysiology. Molecular and Cellular Endocrinology 2019;489:119-125.

Accepted Article

35. Rotroff DM, Pijut SS, Marvel SW, Jack JR, Havener TM, Pujol A, Schluter A, et al. Genetic Variants in HSD17B3, SMAD3, and IPO11 Impact Circulating Lipids in Response to Fenofibrate in Individuals With Type 2 Diabetes. Clinical pharmacology and therapeutics 2018;103:712-721.

36. Su W, Wang Y, Jia X, Wu W, Li L, Tian X, Li S, et al. Comparative proteomic study reveals 17β-HSD13 as a pathogenic protein in nonalcoholic fatty liver disease. Proceedings of the National Academy of Sciences 2014;111:11437-11442.

37. Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. Hepatology 2011;53:1883-1894.

38. Singal AG, Manjunath H, Yopp AC, Beg MS, Marrero JA, Gopal P, Waljee AK. The effect of PNPLA3 on fibrosis progression and development of hepatocellular carcinoma: a meta-analysis. Am J Gastroenterol 2014;109:325-334.

39. Zhang L, You W, Zhang H, Peng R, Zhu Q, Yao A, Li X, et al. PNPLA3 polymorphisms (rs738409) and non-alcoholic fatty liver disease risk and related phenotypes: a meta-analysis. J Gastroenterol Hepatol 2015;30:821-829.

40. Chamorro AJ, Torres JL, Miron-Canelo JA, Gonzalez-Sarmiento R, Laso FJ, Marcos M. Systematic review with meta-analysis: the I148M variant of patatin-like phospholipase domain-containing 3 gene (PNPLA3) is significantly associated with alcoholic liver cirrhosis. Aliment Pharmacol Ther 2014;40:571-581.

41. DiStefano JK, Kingsley C, Craig Wood G, Chu X, Argyropoulos G, Still CD, Done SC, et al. Genome-wide analysis of hepatic lipid content in extreme obesity. Acta Diabetol 2015;52:373-382.

Accepted Article

42. Whitfield JB, Zhu G, Madden PAF, Montgomery GW, Heath AC, Martin NG. Biomarker and Genomic Risk Factors for Liver Function Test Abnormality in Hazardous Drinkers. Alcohol Clin Exp Res 2019;43:473-482.

43. Li Y, Liu S, Gao Y, Ma H, Zhan S, Yang Y, Xin Y, et al. Association of TM6SF2 rs58542926 gene polymorphism with the risk of non-alcoholic fatty liver disease and colorectal adenoma in Chinese Han population. BMC Biochemistry 2019;20:3.

44. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, Vogt TF, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nature genetics 2014;46:352-356.

45. Liu Y-L, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JBS, Allison MED, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nature Communications 2014;5:4309.

46. Liu Z, Que S, Zhou L, Zheng S, Romeo S, Mardinoglu A, Valenti L. The effect of the TM6SF2 E167K variant on liver steatosis and fibrosis in patients with chronic hepatitis C: a meta-analysis. Scientific Reports 2017;7:9273.

47. Luukkonen PK, Zhou Y, Nidhina Haridas PA, Dwivedi OP, Hyötyläinen T, Ali A, Juuti A, et al. Impaired hepatic lipid synthesis from polyunsaturated fatty acids in TM6SF2 E167K variant carriers with NAFLD. Journal of Hepatology 2017;67:128-136.

48. Romeo S, Sanyal A, Valenti L. Leveraging Human Genetics to Identify Potential New Treatments for Fatty Liver Disease. Cell Metab 2020;31:35-45.

49. Trépo E, Valenti L. Update on NAFLD genetics: From new variants to the clinic. J Hepatol 2020;72:1196-1209.

Accepted Article

Table 1

Genome-wide significant associations from primary GWAS analysis.

Top associations at significant loci from GenomALC GWAS										
Chr	BP	rsID	Type	Gene	Effect allele	Effect allele frequency	N cases/controls	Odds Ratio	L95-U95% C.I.	P-value
4	88,214,144	rs4607179	Intergenic	*HSD17B13*	C	0.24	1128 / 614	0.57	0.48-0.62	$1.09x10^{-10}$
5§	175,894,130	rs374702773	Intronic	*FAF2*	del(T)7	0.49	1128 / 614	0.61	0.51-0.73	$2.56x10^{-08}$
22	44,324,727	rs738409	Missense	*PNPLA3*	G	0.39	1128 / 614	2.19	1.81-2.54	$4.93x10^{-17}$

Top associations at significant loci from meta-analysis of GenomALC, UKB, and Buch et al.										
Chr	BP	rsID	Type	Gene	Effect allele	Effect allele frequency	N cases/controls	Odds Ratio	L95-U95% C.I.	P-value
4	88,230,100	rs10433937	Intronic	*HSD17B13*	G	0.25	2071 / 10444	0.78	0.71-0.84	2.85×10^{-09}
5	175,904,141	rs11134977	Intronic	*FAF2*	C	0.46	2071 / 10444	0.79	0.72-0.85	$1.56x10^{-08}$
14	94,844,947	rs28929474	Missense	*SERPINA1*	T	0.02	2071 / 10444	1.9	1.52-2.38	$1.99x10^{-08}$
19	19,407,718	rs10401969	Intronic	*SUGP1/TM6SF2*	C	0.08	2071 / 10444	1.49	1.31-1.70	$2.40x10^{-09}$

Accepted Article

22	44,340,904	rs2294915	Intronic	*PNPLA3*	T	0.32	2071 / 10444	2.07	1.89-2.27	1.28×10^{-53}

Chr (Chromosome), BP (Base Position), rsID (reference SNP cluster ID), Type (type of base substitution), effect allele (coded allele), effect allele frequency (allele frequency of the effect allele in GenomALC cohort), N cases/controls (number of cases / number of controls), L95-U95% C.I. (lower and upper 95% confidence interval).

§ The significant result for the chromosome 5 locus in GenomALC GWAS was obtained after including minor allele counts for rs738409 and rs4607179 as additional covariates.

Accepted Article

Table 2

Phenotypic values among GenomALC cohort.

	GenomALC Cases	**GenomALC Controls**
N	1128	614
Age[1], mean (SD)	52.3 (8.9)	50.1 (9.8)
Sex, N males (% males)	848 (75.2%)	449 (73.1%)
Diabetes, N yes (% yes)	232 (20.6%)	38 (6.2%)
Body Mass Index (BMI), mean (SD)	26.4 (5.3)	24.5 (4.8)
Years of excessive drinking, mean (SD)	24.0 (11.2)	20.8 (9.1)
Alcohol consumed in grams per day, median (Q1, Q3)	160 (104,240)	199 (135,273)

Accepted Article

Total lifetime alcohol consumed in kilograms, median (Q1, Q3)	1315 (760,2237)	1332 (806,2184)

[1] Age for cases (cirrhosis) is calculated as age at first diagnosis of cirrhosis. Age for controls is age at which they enrolled in this trial. SD stands for standard deviation. N indicates number of individuals. Q1 is first quartile and Q3 is third quartile. GenomALC Cases are individuals with alcohol-related liver cirrhosis. GenomALC Controls are heavy drinkers that do not have known liver alcohol-related liver injury.

Accepted Article

Table 3

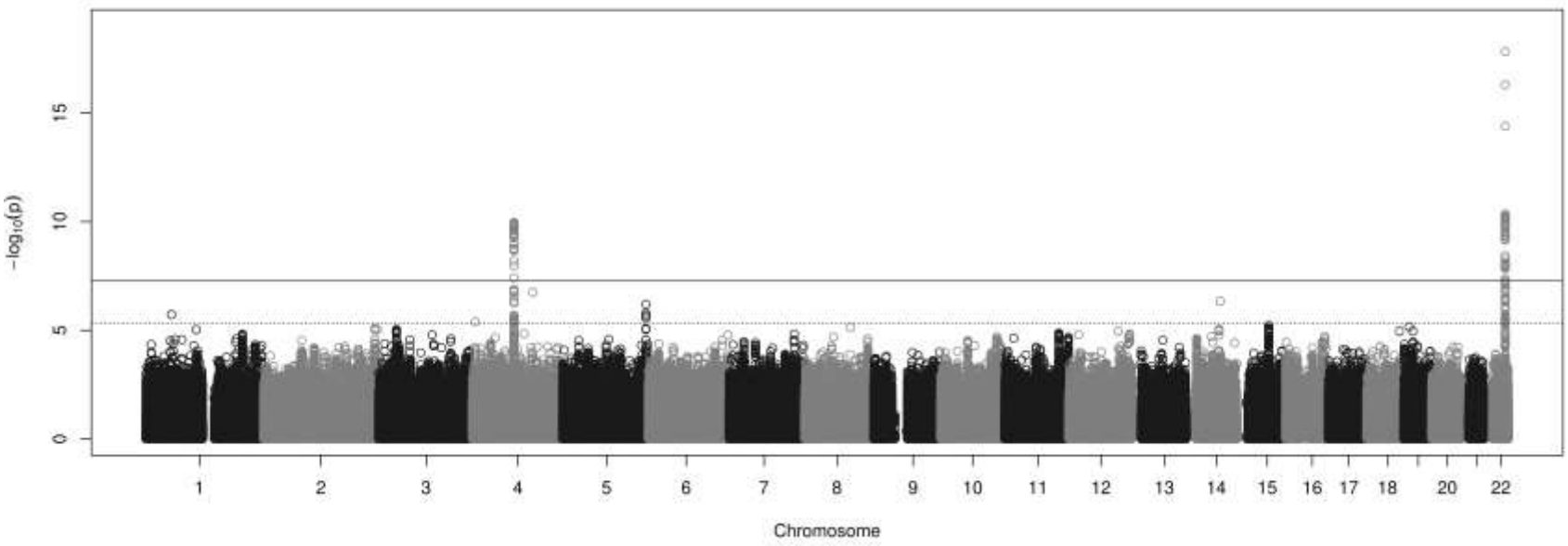
Gene Ontology based enrichment analysis results

	Number of known genes in GO	Number of genes in the list	Expected	Fold Enrichment	+/-	Raw P value	FDR
Lipid droplet (cellular component)	81	3	0.04	70.69	+	9.94E-06	0.020
Lipid droplet organization (biological process)	21	3	0.01	> 100	+	2.15E-07	0.0034

Number of known genes in GO (number of genes with annotated function in Gene Ontology database), Number of genes in the list (number of genes from the input list (also see Supplementary Table 10) that have the same annotated function from Gene Ontology), Expected (expected number of genes in the list from a random set of genes with same size as input list), +/- (increase/decrease in enrichment), Raw P value (Fisher's exact test based p-value), FDR (false discovery rate, adjusted p-value after random permutation)

Accepted Article

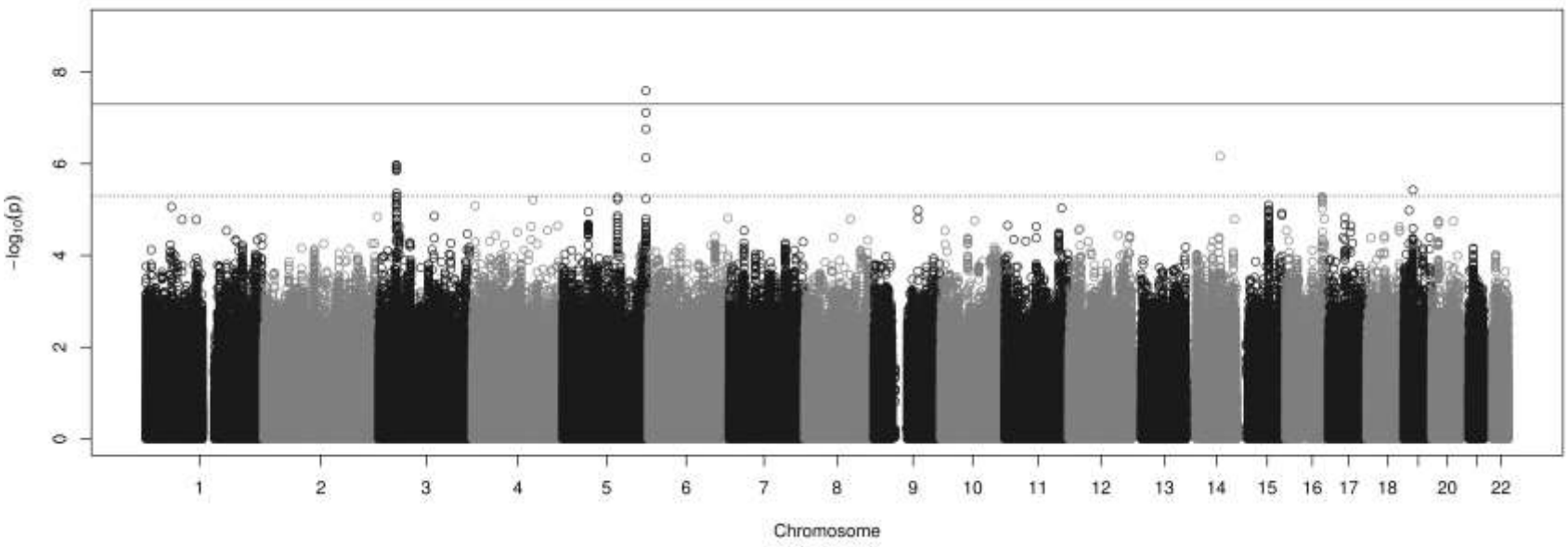
Figure 1



Manhattan plot of primary GWAS analysis of GenomALC. The plot shows significant association of SNPs on chromosomes 4 *(HSD17B13)* and 22 *(PNPLA3)*. Vertical axis shows -log10 transformed p-value of each tested SNP after adjusting for age, sex, diabetes (yes/no), BMI, years of excessive drinking, alcohol consumed in grams per day, total lifetime alcohol consumed in kg, and ten principal components. Horizontal solid grey line shows genome-wide significance level ($p=5x10^{-08}$), dotted grey line shows suggestive level ($p=5x10^{-6}$).

Accepted Article

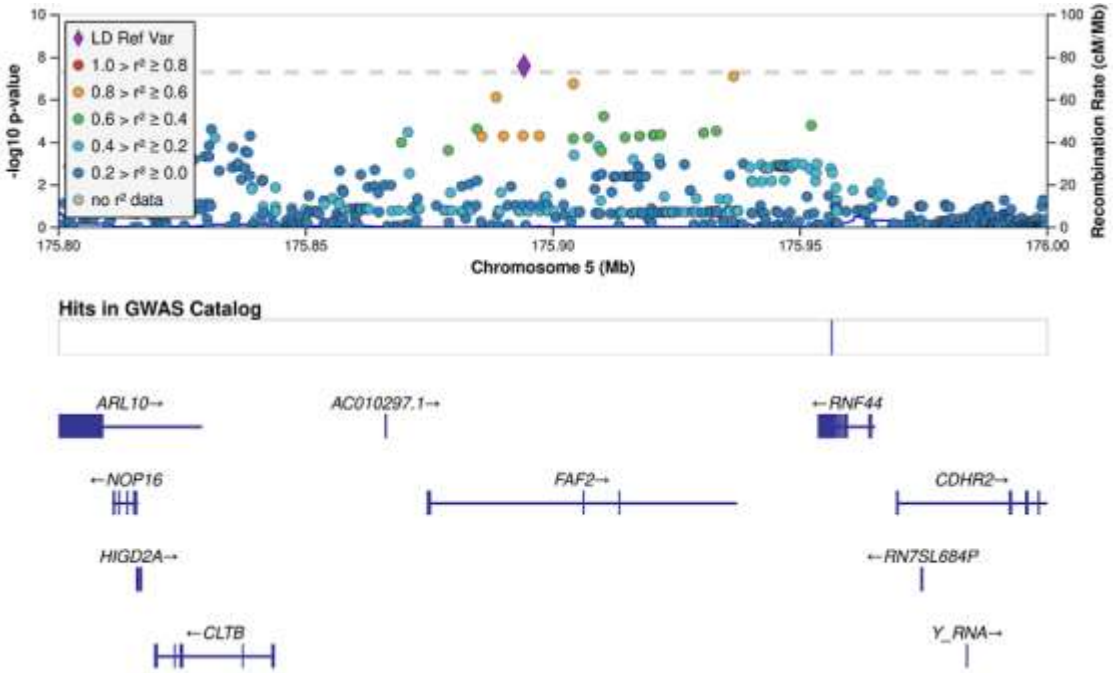
Figure 2



Manhattan plot of secondary GWAS analysis of GenomALC. After accounting for minor allele counts (dosage) for variants in rs4607179 near *HSD17B13* and rs738409 in *PNPLA3*, secondary analysis in GenomALC shows significant association of SNPs on chromosome 5 *(FAF2)*. Vertical axis shows -log10 transformed p-value of each tested SNP after adjusting for rs4607179, rs738409, age, sex, diabetes (yes/no), BMI, years of excessive drinking, alcohol consumed in grams per day, total lifetime alcohol consumed in kg, and ten principal components. Horizontal solid grey line shows genome-wide significance level ($p=5x10^{-08}$), dotted grey line shows suggestive level ($p=5x10^{-6}$).

Accepted Article

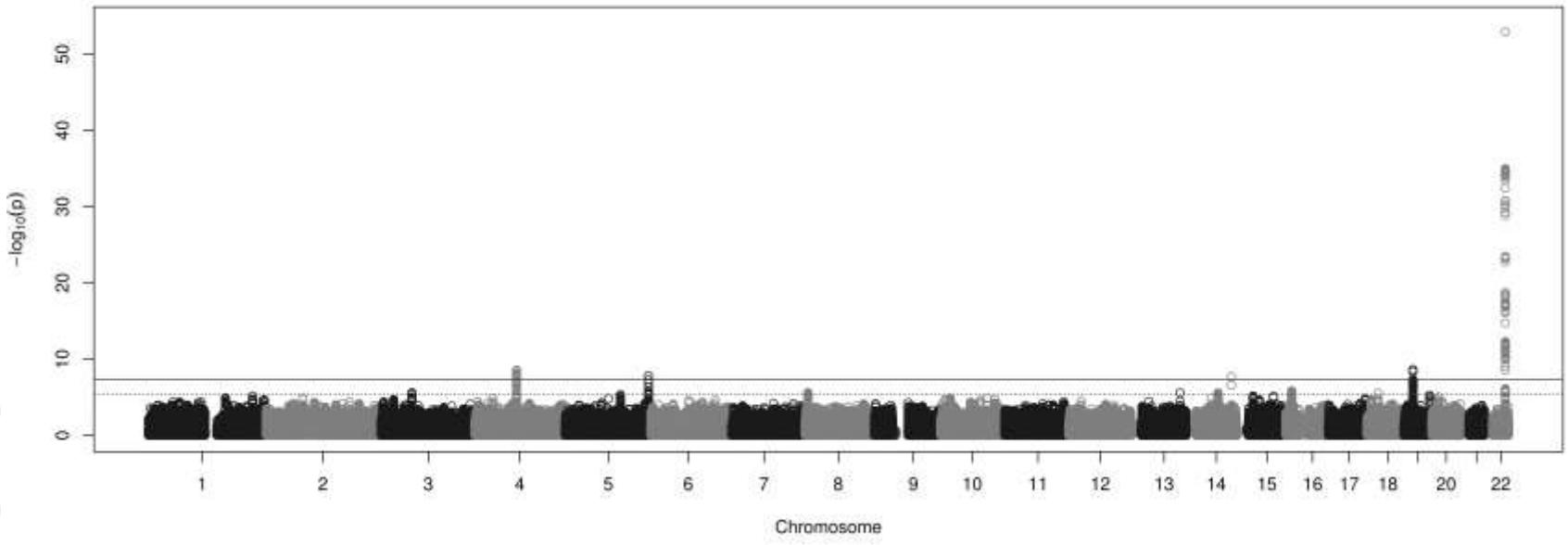
Figure 3



LocusZoom plot of rs374702773 and nearby SNPs on chromosome 5. From conditional analysis of GenomALC data we identified a new locus rs374702773 (purple diamond) on Chr 5 associated with ALC and nearby SNPs in linkage disequilibrium.

Accepted Article

Figure 4



Manhattan plot of meta-analysis of GenomALC, UK Biobank, and Buch et al. Meta-analysis shows significant association of SNPs on chromosomes 4 *(HSD17B13)*, 5 *(FAF2)*, 14 *(SERPINA1)*, 19 *(TM6SF2/SUGP)* and 22 *(PNPLA3)*. Vertical axis shows -log10 transformed p-value of each SNP from meta-analysis including GenomALC and UKB cohort GWASs adjusting for age, sex, diabetes (yes/no), BMI, years of excessive drinking, alcohol consumed in grams per day, total lifetime alcohol consumed in kg, and ten principal components and published results from Buch et al.. Solid grey line shows genome-wide significance level ($p=5x10^{-08}$), dotted grey line shows suggestive level ($p=5x10^{-6}$).